


## Article

# The ‘DEEP’ Landing Error Scoring System

Kim Hébert-Losier <sup>1,\*</sup> , Ivana Hanzlíková <sup>1</sup>, Chen Zheng <sup>2</sup>, Lee Streeter <sup>3</sup> and Michael Mayo <sup>2</sup>

<sup>1</sup> Te Huataki Waiora School of Health, Division of Health, Engineering, Computing and Science, University of Waikato, Tauranga 3116, New Zealand; xhanzlikova@gmail.com

<sup>2</sup> School of Computing and Mathematical Sciences, Division of Health, Engineering, Computing and Science, University of Waikato, Hamilton 3216, New Zealand; zhengchenloot@gmail.com (C.Z.); michael.mayo@waikato.ac.nz (M.M.)

<sup>3</sup> School of Engineering, Division of Health, Engineering, Computing and Science, University of Waikato, Hamilton 3216, New Zealand; lee.streeter@waikato.ac.nz

\* Correspondence: kim.hebert-losier@waikato.ac.nz; Tel.: +64-7-837-9476

Received: 20 December 2019; Accepted: 24 January 2020; Published: 29 January 2020



**Featured Application:** The Landing Error Scoring System, an injury-risk screening tool used in sports to detect high risk of anterior cruciate ligament injury, can be automated using deep-learning-based computer vision on 2D videos combined with machine learning methods. The successful application of this method paves the way for the automatic detection of individuals at high risk of injury using smartphone-based applications and opens doors to addressing other related injury prevention problems.

**Abstract:** The Landing Error Scoring System (LESS) is an injury-risk screening tool used in sports; but scoring is time consuming, clinician-dependent, and generally inaccessible outside of elite sports. Our aim is to evidence that LESS scores can be automated using deep-learning-based computer vision combined with machine learning and compare the accuracy of LESS predictions using different video cropping and machine learning methods. Two-dimensional videos from 320 double-leg drop-jump landings with known LESS scores were analysed in OpenPose. Videos were cropped to key frames manually (clinician) and automatically (computer vision), and 42 kinematic features were extracted. A series of 10 × 10-fold cross-validation experiments were applied on full and balanced datasets to predict LESS scores. Random forest for regression outperformed linear and dummy regression models, yielding the lowest mean absolute error (1.23) and highest correlation ( $r = 0.63$ ) between manual and automated scores. Sensitivity (0.82) and specificity (0.77) were reasonable for risk categorization (high-risk LESS  $\geq 5$  errors). Experiments using either a balanced (versus unbalanced) dataset or manual (versus automated) cropping method did not improve predictions. Further research on the automation would enhance the strength of the agreement between clinical and automated scores beyond its current levels, enabling quasi real-time scoring.

**Keywords:** anterior cruciate ligament; automation; drop jump; injury risk; deep learning; machine learning; movement screen; OpenPose

## 1. Introduction

Lower-extremity injuries due to physical activities have devastating short-term and long-term consequences to the health and wellbeing of individuals [1,2] and burden societies worldwide [3,4]. Non-contact injuries account for approximately 20% of injuries in game situations and 37% of injuries in training situations [5]. Non-contact injuries in sport and recreation are the ones of most practical interest to coaches and clinicians as preventable through neuromuscular training programs [6].

The mechanism of non-contact lower-extremity injuries and their underlying risk factors have been linked with ‘risky’ movement patterns [7,8], such as knee valgus and stiff landings. 3D motion analysis systems, which provide gold-standard measures for the objective quantification of human motion noninvasively, can readily identify altered movement patterns and biomechanical control. However, conventional 3D motion analysis using infrared systems requires a considerable financial outlay and an expert-user, in addition to time and space to perform the analysis. These constraints limit its practical application and use for large-scale screening of injury risk factors in physically active individuals.

As a countermeasure and to reduce technological requirements, various clinician-led movement screens have been developed [9]. Even though these clinician-led screens reduce the financial costs and space requirements compared to 3D motion analysis, they nonetheless require expert clinicians and dedicated time for testing and scoring, limiting their widespread use. For instance, the Functional Movement Screen™ takes 12 to 15 min and the Tuck jump assessment takes 12 min to administer and score for one individual [9].

The Landing Error Scoring System (LESS) is one movement screen with demonstrated reliability [10,11] and validity [11,12]. Clinicians evaluate 2D video recordings from three double-leg drop-jump landing tasks per individual to detect ‘movement errors’ linked to non-contact anterior cruciate ligament (ACL) and other lower-extremity injury mechanisms [10]. The LESS consists of 17 items (Table 1), with the total number of possible errors ranging from 0 (best) to 17 (worst). Greater scores hence indicate more movement errors, poorer landing biomechanics, and greater relative risk of sustaining non-contact lower-extremity injuries. In a prospective study, Padua et al. [12] determined that scoring 5 or more errors on the LESS was associated with a 10.7 times greater relative risk of sustaining a non-contact ACL injury in youth soccer players (sensitivity 0.86, specificity 0.64). The total testing time (including set up) takes ~5 min with 3 to 4 min for a trained rater to score the three drop-jump landing trials of one individual once downloaded to a computer [10].

A few of the drawbacks of the LESS is the subjective nature of the assessment, requirement for an expert-rater, and need to view videos at a later stage [13,14]. In recent years, researchers have striven to automate the LESS to streamline the process using depth sensor cameras [13,15]. Dar, Yehiel, and Cale’ Benzoor [13] introduced the PhysiMax system (PhysiMax Technologies Ltd., Tel Aviv, Israel) to automate LESS scoring using a personal computer, 3D Microsoft Kinect, and motion analysis software that requires limited clinical input. Their results indicated high consensus between clinician and PhysiMax LESS scores (intra-class correlation, ICC = 0.80, mean absolute difference 1.13 errors), although the clinician manually inputted the overall impression item (no. 17, Table 1). Despite the automated quantification of the LESS using markerless motion capture using depth cameras provides time-cost saving benefits, there are still additional hardware-software expenditures to consider.

Deep-learning-based computer vision technologies enable the automatic identification and quantification of human motion without the need for depth sensor cameras. Numerous such systems are currently being developed. For example, OpenPose [16] is a system enabling real-time multi-person pose estimation in video streams captured by a camera. The system tracks both body pose as well as keypoints associated with joints and anatomical features. The same technology is also being deployed for solving other related problems, such as tracking lab animal motion in laboratory settings [17,18]. In this work, we aim to apply deep-learning techniques to LESS score estimation. Applying these approaches to 2D video recordings would improve the accessibility to end-users and pave the way to smartphone-based applications for injury risk screening. Our aim is to evidence that LESS scores can be automated from 2D videos using deep-learning-based computer vision with machine learning and compare the accuracy of LESS predictions using different video cropping and machine learning methods. Our work substantiates that: LESS automation is possible without the need for 3D motion analysis or depth sensor cameras, random forest leads to more accurate predictions than linear or dummy (ZeroR) regression models, and that cropping method (manual versus automated) does not affect predictions.

**Table 1.** Landing Error Scoring System operational definitions of errors. (Adapted from Padua et al. [10].)

No	Item	Definition of Error
1.	Knee flexion IC	Knee flexion < 30°
2.	Hip flexion IC	Thigh is in line with the trunk (hips not flexed)
3.	Trunk flexion IC	Trunk is vertical or extended at the hips (trunk not flexed)
4.	Ankle plantar flexion IC	Heel-to-toe or flat foot landing
5.	Knee valgus IC	The centre of the patella is medial to the midfoot
6.	Lateral trunk flexion IC	The midline of the trunk is flexed to the left or right
7.	Stance width (wide)	Feet are greater than shoulder width apart
8.	Stance width (narrow)	Feet are less than shoulder width apart
9.	Foot (toe-in)	Foot is externally rotated > 30° between IC and KF <sub>max</sub>
10.	Foot (toe-out)	Foot is internally rotated > 30° between IC and KF <sub>max</sub>
11.	Symmetric foot contact IC	One foot lands before the other One foot lands heel-toe and the other foot lands toe-heel
12.	Knee flexion displacement	Knee flexes < 45° between IC and KF <sub>max</sub>
13.	Hip flexion at KF <sub>max</sub>	Thigh does not flex more on trunk from IC to KF <sub>max</sub>
14.	Trunk flexion at KF <sub>max</sub>	Trunk does not flex more from IC to KF <sub>max</sub>
15.	Knee valgus displacement	At max medial knee position, centre of the patella is medial to the midfoot
16.	Joint displacement	Soft, average, stiff
17.	Overall impression	Excellent, average, poor

Abbreviations: IC, initial contact; KF<sub>max</sub>, maximal knee flexion.

## 2. Materials and Methods

### 2.1. Participants

A sample of 144 individuals (45 males and 99 females) volunteered to participate in this study. Age, height, mass, and body mass index (mean  $\pm$  standard deviation) for males were  $21.0 \pm 5.9$  years (range 17 to 42 years),  $179.1 \pm 7.2$  cm, and  $82.2 \pm 13.6$  kg; and for females were  $17.1 \pm 3.7$  years (range 12 to 31 years),  $169.2 \pm 6.1$  cm, and  $64.8 \pm 9.6$  kg. All participants were involved in physical activity (34% participated in netball, 19% in rugby, 9% in field hockey, 9% in soccer, and 29% in other sports). On average, participants were involved in physical activity four times per week, 6 h a week. Participants had to be free from injury, pain, or any other issue that would limit physical activity participation. Previous injuries were not an exclusion criterion. Participants were recruited via word-of-mouth, research contacts, social media, and emails sent to local sports clubs. The study protocol was approved by our institution's health research ethics committee [HREC(Health)#41] and adhered to the Declaration of Helsinki. All participants and their legal guardian when younger than 16 years of age signed a written informed consent document that explained the potential risks associated with testing prior to participation.

### 2.2. Data Collection

We used the original LESS protocol for testing [10]. Participants jumped horizontally from a 30 cm high box to a line placed at 50% of their body height, and immediately jumped upward for maximal vertical height. We placed an emphasis on jumping off the box with both feet, landing in front of the designated line, jumping as high as possible straight up in the air once they landed from the box, and completing the task in a fluid motion. We did not provide any feedback on participants landing technique unless they were performing the task incorrectly. Participants used their own footwear for testing.

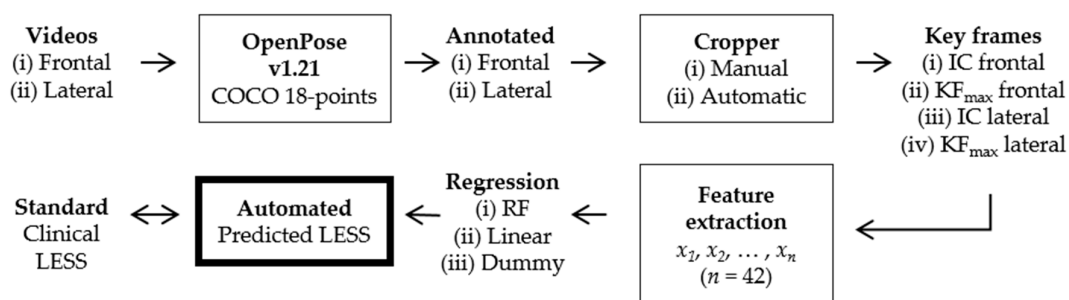
After task instructions and practice jumps for familiarization (typically 1), each participant performed three successful trials of the double-leg drop-jump landing task in front of two standard video cameras capturing at 120 Hz (Sony RX10 II, Sony Corporation, Tokyo, Japan) with an actual focal length of 8.8 to 73.3 mm (35 mm equivalent focal length of 24–200 mm). We mounted the cameras on tripods placed 3.5 m in front of and to the right side of the landing area with a lens-to-floor distance of 1.3 m. We allowed participants to rest until they felt ready to perform the task again to limit fatigue between the three trials. Total testing time was typically 2 min per participant.

### 2.3. Clinical LESS

A qualified physiotherapist who completed over 400 LESS evaluations (IH) replayed the videos using the Kinovea software (version 0.8.15, [www.kinovea.org](http://www.kinovea.org)), identified the two key frames of initial ground contact (IC) and maximal knee flexion ( $KF_{max}$ ), and scored all trials using the 17-item LESS scoring sheet (Table 1). The clinician was blinded to the results from the automated computer-vision scoring. A total of 320 double-leg drop-jump landings from the potential 432 trials (3 jumps  $\times$  144 participants) were retained for analysis because of certain participants not completing three trials, one or both video files being not usable, or a clear misidentification of time events from the automatic cropper described in the following subsection (i.e., more than 100 ms difference with the clinician).

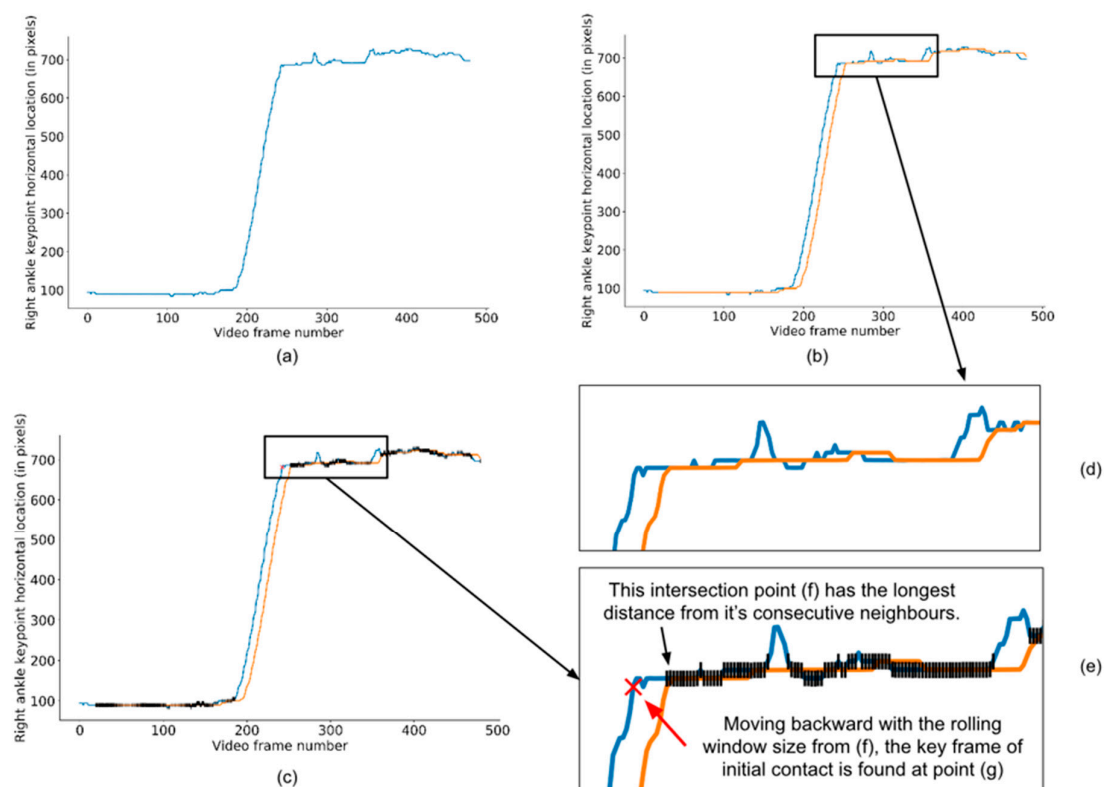
### 2.4. Automated LESS

The LESS score prediction algorithm we developed was a multistage process. Generally, the first stage consisted of processing the videos to detect the IC and  $KF_{max}$  key frames, which involved running the frontal and lateral videos for each jump through OpenPose v1.21 [16], and then using a heuristic method to identify the key frames. Once that stage was complete, we extracted measurements from the key frames to use as features for machine learning. The final stage was the score prediction for the drop-jump landing trial from the features using a machine learning algorithm. The entire process is depicted in Figure 1. We further evaluated the predictive accuracy of the final machine learning stage using cross validation.



**Figure 1.** Flow diagram of data processing leading to comparing ‘gold standard’ clinical LESS scores from an expert rater to ‘automated’ predicted LESS scores from the automation process. *Abbreviations:* IC, initial contact;  $KF_{max}$ , maximal knee flexion; LESS, Landing Error Scoring System; RF, random forest.

In more detail, the algorithm used to detect key frames in the first stage is described in Table 2. The input to the algorithm are the frontal and lateral videos for a single drop-jump landing trial, and the output are cropped versions of the same videos where the first and last frames correspond to the IC and  $KF_{max}$  key frames, respectively.



**Figure 2.** This figure is an example of the original (blue line) plot and rolling window (orange line) plot for the right ankle keypoint of one individual during a drop-jump landing trial taken from the lateral view video. More specifically, (a) the blue line depicts the distance of the right ankle to the left boarder (y-axis) in each video frame (x-axis); (b) the orange line is the 20-frame rolling median of the original blue line; (c) the black bars indicate the intersections of the two lines, whereas the red dotted line represents the distance between two consecutive intersection points. Figure (d) is a zoomed-in view of the intersections around the initial contact key frame. Figure (e) highlights the points (f) and (g) as the initial contact key frame on the rolling window plot and original plot, respectively.

The basic method is to track the location of the ankles (using OpenPose and COCO 18-points model [16]) across the frames to detect the frame in which landing occurs based on the original and rolling window plots (Figure 2), and additionally to track the body and knee keypoints so that the ankle/knee/body angle can be calculated and used to identify the point of maximum knee flexion. Once these two points are identified in both videos, then the frames before and after the key frames are cropped away. This stage generally reduces the length of the original videos from several seconds down to less than 250 ms.

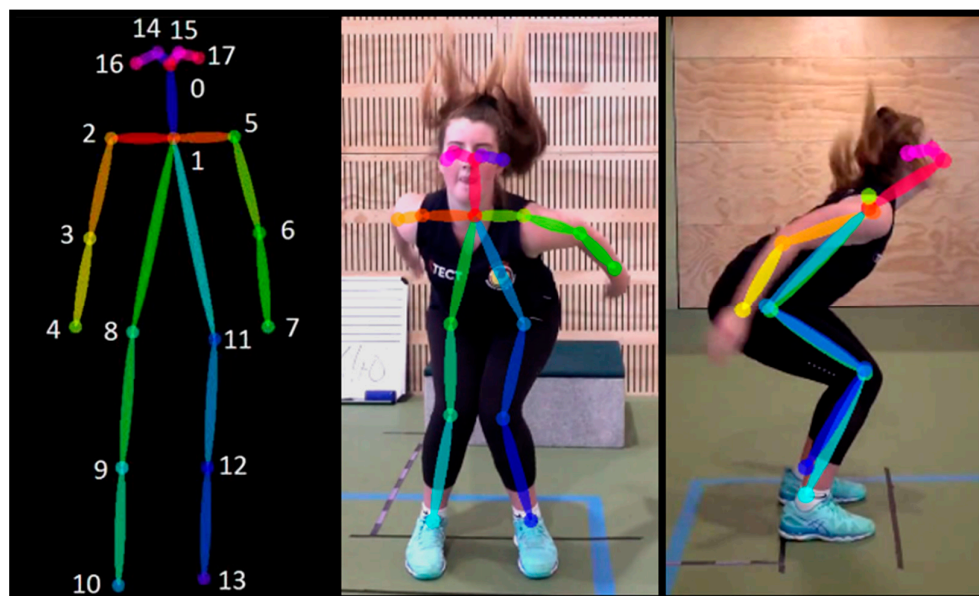
Once cropping is complete, two videos in which the first frame corresponds to IC and the last frame corresponds to  $KF_{max}$  pass to the second stage. In the second stage of processing, features are extracted from both videos and merged into a single 'example' that will be used for machine learning. A total of 42 kinematic features from the two key frames in each video were generated. The features are a mixture of angles between specific OpenPose keypoints (shown in Figure 3) and ratio between distances. The specific features are listed in Table 3. A total of six angles were extracted from all four key frames with an additional eight features (mixture of angles, distances, and distance ratios) being extracted from the two frontal key frames only, for a total of 40 measurements. Two further features, being the length in frames of the cropped frontal and lateral videos, were also included.



**Table 2.** Algorithm used to detect key frames from the two input videos.

Step	Description
	Input: F, Frontal view video; L, Lateral view video
1.	Obtain the body part keypoints in each frame in both F and L videos using OpenPose
2.	Impute keypoint positions using linear interpolation when not recognized by OpenPose
3.	Find F key frames IC and KF <sub>max</sub>
3.1.	Based on the coordinates of the left and right ankle (both visible in F), find the intersections of the original and rolling window plots <sup>a</sup> for each ankle
3.2.	Calculate the distances between each consecutive intersection point pairs
3.3.	Find the first point of the pair of intersection points with the longest distances for each ankle
3.4.	Identify the first point of the pair of intersection points that has the lowest <i>x</i> value (i.e., number of frames) as IC
3.5.	Based on the coordinates of the body keypoint, find the intersections of the original and rolling window plots <sup>a</sup>
3.6.	Calculate the distances between each consecutive intersection point pairs
3.7.	Identify the first point of the pair of intersection points with the longest distances as KF <sub>max</sub>
4.	Find L key frames IC and KF <sub>max</sub>
4.1.	Based on the coordinates of the individual's right ankle (which is closest to the camera L), find the intersections of the original and rolling window plots
4.2.	Calculate the distances between each consecutive intersection point pairs
4.3.	Identify the first point of the pair of intersection points with the longest distances as IC
4.4.	Based on the coordinates of the body keypoint, find the intersections of the original and rolling window <sup>a</sup> plots
4.5.	Calculate the distances between each consecutive intersection point pairs
4.6.	Identify the first point of the pair of intersection points with the longest distances with upper/positive trend as KF <sub>max</sub>
5.	Crop the videos (F and L) according to IC and KF <sub>max</sub> key frames.
	Output: F', cropped version of frontal view video; L, cropped version of lateral view video

Notes. <sup>a</sup> Rolling window plot, plot of median values from a rolling 20-frame window. See Figure 2. Abbreviations. F, Frontal view video; IC, initial contact; L, Lateral view video KF<sub>max</sub>, maximal knee flexion.



**Figure 3.** OpenPose's COCO 18-points model keypoint positions (left image) [16] and example of a frontal (middle image) and lateral (right image) view processed video at the maximal knee flexion key frame.

**Table 3.** Measurements extracted from key frames and used as kinematic features.

Key Frames and Views	Measurement (OpenPose Numbers <sup>a</sup> )	Kinematic Features
<b>All four key frames</b> (two frontal key frames and two lateral key frames)	Angle (8,9,10)	Right knee angle
	Angle (9,8,1)	Right hip angle
	Angle (2,1,8)	Right trunk angle
	Angle (3,2,1)	Right shoulder angle
	Angle (4,3,2)	Right elbow angle
	Angle (2,1,0)	Right neck angle
<b>Two key frames</b> (two frontal key frames)	Angle (11,12,13)	Left knee angle
	Angle (1,11,12)	Left hip angle
	Angle (11,1,5)	Left trunk angle
	Angle (1,5,6)	Left shoulder angle
	Angle (5,6,7)	Left elbow angle
	Distance (9,12)	Knee distance
	Distance (2,5)	Shoulder distance
	Distance (9,12)/Distance (2,5)	Knee distance/Shoulder distance

Notes. Key frames are: (i) initial contact, (ii) maximal knee flexion. <sup>a</sup> Refer to Figure 3 for keypoints.

Following feature extraction, we then used a machine learning algorithm to predict the LESS score associated with the drop-jump landing videos. To evaluate the predictive effectiveness of the various machine learning algorithms, we generated features for all 320 drop-jump landings in the dataset using the approach described above. It was also noticed that the distribution of the LESS scores in the dataset was imbalanced, with the majority of LESS scores falling in the range 4–6. Given that unbalanced datasets can potentially affect the accuracy of machine learning techniques, we additionally generated a balanced version of the dataset consisting of 153 drop-jump landing trials with at most 20 trials per LESS score. All evaluations of machine learning techniques were applied to both datasets.

The machine learning techniques chosen to be evaluated were random forest regression, because it is a state-of-the-art machine learning approach and generally performs well ‘out of the box’ on most problems in practice; and linear regression, which is a widely understood linear modelling technique. Unlike random forest regression, linear regression produces an interpretable model, but it has the disadvantage of being unable to model interactions between features. Given that the full dataset was imbalanced, we also evaluated a dummy regressor (ZeroR) that simply predicts the mean LESS score from the training data. For the original dataset, this method was expected to have reasonably high accuracy, but lower accuracy for the balanced dataset. All machine learning methods implemented were available in WEKA 3.8.0 [19], and returned floating point numbers (i.e., decimals) that added granularity to the data.

## 2.5. Statistical Method

As noted in Section 2.3, 320 double-leg drop-jump landings were analysed. A series of 10 × 10-fold cross validation experiments were applied on full (320 videos) and balanced (153 videos, ≤ 20 videos per LESS score) to predict the scores using random forest for regression, linear regression, and dummy regression (ZeroR) models in WEKA [20]. To assess the effectiveness of the automated cropping algorithm in the context of the overall system, we additionally ran the entire pipeline with crops generated by the clinician. Mean absolute error and Pearson correlation coefficient (*r*) were calculated to assess the accuracy of the predictions. Predictions were then converted to a binary category and sensitivity-specificity for categorising individuals at high risk of non-contact ACL injury (LESS ≥ 5 errors [12]) were assessed for each method. The outcomes of the models were compared using paired corrected *t*-tests in WEKA [20], and the timestamps of the key frames IC and KF<sub>max</sub> respectively compared between manual (clinician) and automated (OpenPose) cropping methods using unpaired *t*-tests assuming homoscedasticity. Since the LESS score was treated as a regression problem, actual (clinical LESS) versus predicted (automated LESS) and Bland-Altman [21] plots were used to allow for a visual inspection of the models. Statistical significance was set at *p* ≤ 0.05.

### 3. Results

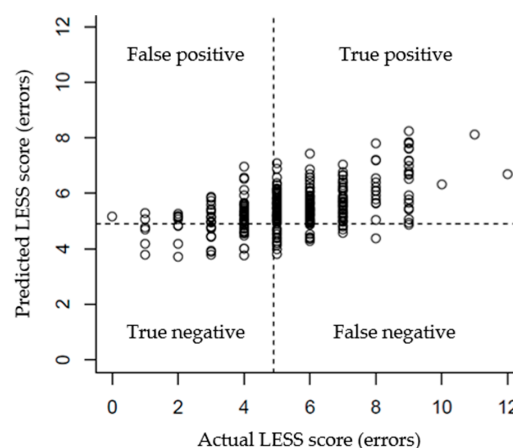
The mean LESS score from the 320 drop-jump landings was  $5.5 \pm 1.8$  errors (range 0 to 12 errors) as rated by the clinician. The absolute time difference between manually identified IC and  $KF_{\max}$  was  $26.5 \pm 17.0$  ( $p = 0.484$ ) and  $32.8 \pm 18.0$  ms ( $p = 0.445$ ) for the frontal videos, and  $53.5 \pm 16.2$  ( $p = 0.125$ ) and  $20.8 \pm 16.3$  ms ( $p = 0.827$ ) for the sagittal videos.

Random forest yielded the lowest mean absolute error (1.23) and greatest correlation ( $r = 0.63$ ) between actual and predicted scores based on results from the cross validation experiments (Table 4). Sensitivity (0.82) and specificity (0.77) were reasonable for high (LESS  $\geq 5$  errors) and low (LESS  $< 5$  errors) injury risk categorisation. Experiments using a balanced (versus unbalanced) dataset or manually (versus automated) cropping methods did not improve predictions. An actual versus predicted plot from the random forest regression is depicted in Figure 4, and two Bland-Altman plots on the same dataset in Figure 5. Note that both conventional (mean difference  $\pm 1.96$  standard deviation) and regression-based (regressed difference between methods on the mean of the two methods  $\pm 2.46$  standard deviation of the residual) Bland-Altman plots were generated given the non-uniform differences in mean [21].

**Table 4.** Results from machine learning experiments.

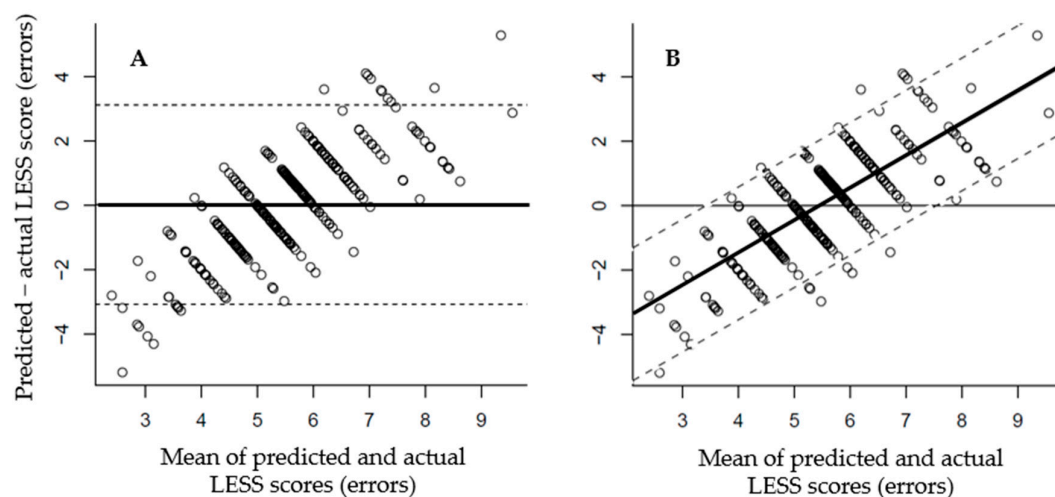
Cropper	Dataset	Mean Absolute Error (n Errors)			Correlation ( $r$ )		
		RF	Linear	Dummy	RF	Linear	Dummy
(i) Manual	(i) Full	$1.23 \pm 0.18$	$1.39 \pm 0.20^*$	$1.44 \pm 0.20^*$	$0.52 \pm 0.15$	$0.39 \pm 0.14^*$	$0.0 \pm 0.0^*$
	(ii) Balanced	$1.57 \pm 0.27$	$1.90 \pm 0.61$	$2.08 \pm 0.34^*$	$0.60 \pm 0.15$	$0.48 \pm 0.21^*$	$0.0 \pm 0.0^*$
(ii) Automatic	(i) Full	$1.23 \pm 0.18$	$1.32 \pm 0.20^*$	$1.44 \pm 0.20^*$	$0.53 \pm 0.15$	$0.44 \pm 0.16^*$	$0.0 \pm 0.0^*$
	(ii) Balanced	$1.56 \pm 0.29$	$1.63 \pm 0.32$	$2.08 \pm 0.32^*$	$0.63 \pm 0.17$	$0.51 \pm 0.20^*$	$0.0 \pm 0.0^*$
Cropper	Dataset	Sensitivity <sup>a</sup>			Specificity <sup>a</sup>		
		RF	Linear	Dummy	RF	Linear	Dummy
(i) Manual	(i) Full	$0.80 \pm 0.09$	$0.75 \pm 0.09$	$1.0 \pm 0.0^*$	$0.50 \pm 0.18$	$0.51 \pm 0.18$	$0.0 \pm 0.0^*$
	(ii) Balanced	$0.77 \pm 0.13$	$0.73 \pm 0.13$	$1.0 \pm 0.0^*$	$0.73 \pm 0.18$	$0.63 \pm 0.21$	$0.0 \pm 0.0^*$
(ii) Automatic	(i) Full	$0.82 \pm 0.07$	$0.77 \pm 0.09^*$	$1.0 \pm 0.0^*$	$0.52 \pm 0.19$	$0.52 \pm 0.18$	$0.0 \pm 0.0^*$
	(ii) Balanced	$0.76 \pm 0.15$	$0.77 \pm 0.13$	$1.0 \pm 0.0^*$	$0.77 \pm 0.19$	$0.70 \pm 0.21$	$0.0 \pm 0.0^*$

Notes. Values are means  $\pm$  standard deviations. Abbreviations. RF, random forest. \* Significant difference versus random forest ( $p \leq 0.05$ ) using paired-corrected t-tests. <sup>a</sup> Categorising high (LESS  $\geq 5$  errors) and low (LESS  $< 5$  errors) injury risk individuals [12].



**Figure 4.** Actual (clinical) versus predicted (automated) LESS score plots from the random forest regression using full dataset ( $n = 320$ ) and automatic cropping method. Dashed lines represent the 5-error threshold that defines high risk of injury (i.e., scoring 5 or more errors during LESS has been associated with a 10.7 times greater relative risk of sustaining a non-contact anterior cruciate ligament injury [12]). Note that the clinical scores are integers and predicted scores are decimals, which adds granularity. Abbreviations: LESS, Landing Error Scoring System.





**Figure 5.** Bland-Altman [21] plots depicting the difference in predicted (automated) and actual (clinical) LESS scores versus the mean scores with (A) conventional 95% limits of agreement (mean difference  $\pm 1.96$  standard deviation), and (B) regression-based limits of agreement (regressed difference between methods on the mean of the two methods  $\pm 2.46$  standard deviation of the residual).

#### 4. Discussion

The use of the LESS to assess injury risk is common in sport science and clinical practice [9,22], but scoring is time consuming, clinician-dependent, and generally inaccessible for large-scale screening outside of elite sports. This study provides evidence that the LESS can be automated using deep-learning-based computer vision combined with machine learning methods without the need for 3D motion analysis or depth sensor cameras. A clear benefit of automating LESS scoring is immediate feedback to end-users. The successful application of this method paves the way for the automatic detection of individuals at high risk of injury using smartphone-based applications of LESS videos (Video S1: <https://youtu.be/q1wiGt4K8MU>).

The characteristics of an ideal injury risk screening tool are good reliability, validity, and predictive value for injury incidence. In practical or field settings, an ideal screening method is easy to administer without an expert, and has minimal financial, spatial, and temporal requirements. Ideally, the screening tool provides immediate results and is accessible to everyone, from the recreational to elite athlete, as well as novice to expert rater. Overall, the LESS responds to most of these stated requirements. The test demonstrates acceptable reliability and validity [10,11,23], as well as predictive value for non-contact ACL injury using a threshold of 5 errors [12]. The inter-rater reliability of the total LESS score is good to excellent, with ICC ranging from 0.83 to 0.92 [10,11,23] and typical errors at 0.71 LESS errors [10]. The results from the current study indicate that the typical errors from the automated processing and scoring of the LESS through computer vision when applying the random forest model (Table 2) are less than half an error greater than scores taken from two expert clinicians. In fact, certain individual LESS items yield suboptimal psychometric properties between raters and 3D motion analysis [23]. More specifically, no significant agreement between raters was found for knee and trunk flexion at IC, and poor agreement between rater and 3D motion capture analysis was found for knee flexion at IC, lateral trunk flexion at IC, and symmetric foot contact at IC [23]. As such, a certain level of disagreement between clinical ratings and computerised ratings is expected.

As seen in Figures 4 and 5, the estimated error is not uniform across the range of LESS scores, but depends on the target. For example, trials with a low actual LESS score tend to have a positive error (the prediction is an overestimation) and trials with a higher actual LESS score tend to have a negative error (the prediction is an underestimation). If these biases stemmed from the over representation of the mid-range LESS values (i.e., majority of LESS scores falling in the range 4–6), the balanced dataset should have provided more accurate predictions, which was not the case. It might be possible

to attempt correcting predictions to improve accuracy in future work using probability calibration methods, such as Platt Scaling and Isotonic Regression. The large errors in LESS score predictions were attributed to inaccurate foot and IC key frame detection. The newest body model in OpenPose (Body 25) contains 25 points, including coordinates that define the feet and enable computations of angles at the ankles [24]. Improving the LESS score automation relies on either refining body part detection or training a new system specifically to solve this problem.

In previous research, depth sensor technology has been used to automate LESS scoring [13,15]. Comparisons between automated and expert clinicians indicate a mean difference of 1.20 errors [15], mean absolute difference of 1.13 errors [13], intra-class correlation of 0.80 [16], and percentage agreement of the individual items ranging from 55–100% [13,15]. These research findings are comparable to our lowest mean absolute error (1.23), greatest correlation ( $r = 0.63$ ), and agreement in risk classification (sensitivity 0.82, specificity 0.77) between actual and predicted scores from the cross validation experiments using random forest regression. In contrast to the PhysiMax system [13,15], our approach did not require the clinician to add the overall impression manually (no. 17 in Table 1) given that the LESS items were not scored one-by-one. Although the lack of individual-item scores might be perceived as a limitation of the deep LESS approach; no subjective rating from the clinician or hardware other than a handheld camera or smart portable device are required. Furthermore, only the final LESS score has shown predictive value in terms of injury risk [12]; hence, the individual items are of lesser clinical value.

The better accuracy achieved by random forest can be explained by the fact that the features (angles, distances, and ratios) are likely correlated and related in a non-linear manner. Decision tree ensembles in general are better able to cope with correlated variables and model non-linear patterns [25]. Linear regression, on the other hand, achieves optimal results when the predictor variables are independent and do not interact. We also foresee a possibility of processing the raw video images themselves and attempting direct deep learning-based classification with minimal pre-processing. Such an approach would obviate the need to use OpenPose or a similar pose-tracking tool. However, taking such an approach would be challenging because of the lack of training data relative to size of datasets usually used to train deep image recognisers. Another significant disadvantage of the proposed approach is that deep learning needs GPU-based acceleration hardware, and is therefore currently unable to process videos independently on consumer smartphones. That said; the rapidly increasing computational power of consumer smartphones and the current trend in research of compressing deep models [26] so that they run efficiently on mobile devices should solve this problem in the next few years.

One of the main concerns in clinical screening tools are their subjective nature and reliance on visual observations to estimate angles, which are challenging to quantify accurately [27,28]. During the LESS, a small kinematic difference (e.g., knee angle  $29^\circ$ , 1—error present; knee angle  $30^\circ$ , 0—error absent) can result in poor agreement between raters and between clinical LESS scores and motion capture scores. Recent technological advances have allowed the more objective quantification of human motion using wearable technology [29,30]. Inertial measurement units are able to measure linear and angular motion of individual body segments and centre of mass, and are proposed as more accurate means of identifying risky movement patterns than through visual observations [31]. Although inertial measurement units are relatively inexpensive; they are not commonly used in clinical environments and an expert is still needed to process and interpret data signals. The automated scoring process here developed using standard video recordings offers an alternative solution that can possibly improve consistency of LESS ratings, removing the subjective interpretation of the task. Moving forward, the reliability of deep LESS scores, validity of OpenPose derived data during the dynamic double-leg drop-landing task, and predictive ability of the method need empirical support.

An indisputable advantage of automated scoring using deep-learning-based computer vision combined with machine learning methods or markerless methods from depth sensor cameras is immediate results and feedback to patients, athletes, coaches, or healthcare professionals.

Our developed method that automates LESS scores provides a viable solution to decreasing scoring time, increasing accessibility to non-expert raters, and delivering immediate results without any additional expenditure other than conventional video recordings. Conventional 2D video recordings are adequate for quantifying kinematics [32–34] and are readily accessible through tablets or smartphones. The successful application of this method would pave the way for the automatic detection of individuals at high risk of injury using smartphone-based applications of LESS and 2D video footage (Video S1: <https://youtu.be/q1wiGt4K8MU>). Other than expediting mass injury risk screening initiatives in youth or team sports, LESS automation could be a valuable and convenient tool to track injury risk factors over time and to assess the effectiveness of intervention programs at improving landing mechanics (Video S2: [https://youtu.be/Ve\\_QJu0fuLs](https://youtu.be/Ve_QJu0fuLs)). The proposed method could be extended to other injury risk screening methods based on 2D camera recordings to decrease manual labour and time required for screening initiatives; e.g., the Cutting Movement Assessment Scale [35] and Tuck jump assessment [36].

This preliminary investigation provides evidence that it is feasible to automate the LESS from 2D video recordings alone. Further research could lead to improved automation outcomes and enhance the strength of the agreement between clinical and automated LESS scores beyond its current levels. The newest body model in OpenPose (Body 25) contains 25 points, including coordinates that define the feet and enable computations of angles at the ankles [24]. Although the timestamped IC key frame in frontal and sagittal videos were comparable between the clinician and scripted process (mean difference: 32.8 ms,  $p = 0.445$  and  $20.8 \pm 16.3$  ms,  $p = 0.827$ ), using the foot coordinates rather than ankle and body coordinates would certainly enhance precision. A number of videos from the available dataset were not used because of a clear misidentification of time events from the automatic cropper (i.e., more than 100 ms). We were unable to determine the reason underlying the mislabelling of these videos upon visual inspection. We speculate that rerunning the current experiment using the COCO + Foot model might lead to the correct identification of key events in a greater number of our database videos, increasing the number of eligible videos for analysis. The increased number of coordinates from the 25-point Body rather than 18-point COCO model would also allow us to extract a greater number of features from the processed videos and use these as input in the subsequent regression experiments.

## 5. Conclusions

We provide evidence that the Landing Error Scoring System (LESS)—an injury-risk screening tool—can be automated using deep-learning-based computer vision combined with machine learning methods. Further research on the automation would enhance the strength of the agreement between clinical (gold standard) and automated (predicted) LESS scores, and risk classification beyond its current levels. Automation of the LESS using standard 2D recordings would facilitate mass injury-risk screening initiatives with quasi real-time feedback, without the need of depth cameras or expert clinicians. The successful application of this method would pave the way for the automatic detection of individuals at high risk of injury using smartphone-based applications of LESS and 2D video footage (Video S1: <https://youtu.be/q1wiGt4K8MU>), increasing accessibility of injury-risk assessment methods beyond elite athletes and removing depth-sensor camera requirements. It may also open doors to other related injury prevention problems. Future work includes updating the framework using the newest body model in OpenPose (Body 25) to extract a greater number of features and more accurately detect key frames.

**Supplementary Materials:** The following are available online: <https://youtu.be/q1wiGt4K8MU>. Video S1: LESS demonstration. [https://youtu.be/Ve\\_QJu0fuLs](https://youtu.be/Ve_QJu0fuLs). Video S2: The ‘DEEP’ Landing Error Scoring System.

**Author Contributions:** Conceptualization, K.H.-L.; methodology, K.H.-L.; formal analysis, K.H.-L., L.S., M.M.; investigation, K.H.-L., I.H.; data curation, I.H., C.Z.; writing—original draft preparation, K.H.-L., I.H.; writing—review and editing, K.H.-L., I.H., C.Z., L.S., M.M.; supervision, K.H.-L., M.M.; project administration, K.H.-L.; funding acquisition, K.H.-L., L.S., M.M. Authorship must be limited to those who have contributed substantially to the work reported. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a University of Waikato Strategic Investment Fund 2018 Medium research Grant.

**Acknowledgments:** We would like to acknowledge Ruili Wang for expert advice and Christopher Martyn Beaven for research support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hébert-Losier, K.; Pini, A.; Vantini, S.; Strandberg, J.; Abramowicz, K.; Schelin, L.; Häger, C.K. One-leg hop kinematics 20years following anterior cruciate ligament rupture: Data revisited using functional data analysis. *Clin. Biomech.* **2015**, *30*, 1153–1161. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Hébert-Losier, K.; Schelin, L.; Tengman, E.; Strong, A.; Hager, C.K. Curve analyses reveal altered knee, hip, and trunk kinematics during drop-jumps long after anterior cruciate ligament rupture. *Knee* **2018**, *25*, 226–239. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Mather, R.C., 3rd; Hettrich, C.M.; Dunn, W.R.; Cole, B.J.; Bach, B.R., Jr.; Huston, L.J.; Reinke, E.K.; Spindler, K.P. Cost-effectiveness analysis of early reconstruction versus rehabilitation and delayed reconstruction for Anterior Cruciate Ligament tears. *Am. J. Sports Med.* **2014**, *42*, 1583–1591. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Sutherland, K.; Clatworthy, M.; Fulcher, M.; Chang, K.; Young, S.W. Marked increase in the incidence of anterior cruciate ligament reconstructions in young females in New Zealand. *ANZ J. Surg.* **2019**, *89*, 1151–1155. [\[CrossRef\]](#)
5. Hootman, J.M.; Dick, R.; Agel, J. Epidemiology of collegiate injuries for 15 sports: Summary and recommendations for injury prevention initiatives. *J. Athl. Train.* **2007**, *42*, 311–319.
6. Al Attar, W.S.A.; Alshehri, M.A. A meta-analysis of meta-analyses of the effectiveness of FIFA injury prevention programs in soccer. *Scand. J. Med. Sci. Sports* **2019**, *29*, 1846–1855. [\[CrossRef\]](#)
7. Hewett, T.E.; Myer, G.D.; Ford, K.R. Anterior cruciate ligament injuries in female athletes: Part 1, mechanisms and risk factors. *Am. J. Sports Med.* **2006**, *34*, 299–311. [\[CrossRef\]](#)
8. Leppänen, M.; Pasanen, K.; Kujala, U.M.; Vasankari, T.; Kannus, P.; Ayramo, S.; Krosshaug, T.; Bahr, R.; Avela, J.; Perttunen, J.; et al. Stiff landings are associated with increased ACL injury risk in young female basketball and floorball players. *Am. J. Sports Med.* **2017**, *45*, 386–393. [\[CrossRef\]](#)
9. Chimera, N.J.; Warren, M. Use of clinical movement screening tests to predict injury in sport. *World J. Orthop.* **2016**, *7*, 202–217. [\[CrossRef\]](#)
10. Padua, D.A.; Marshall, S.W.; Boling, M.C.; Thigpen, C.A.; Garrett, W.E., Jr.; Beutler, A.I. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. *Am. J. Sports Med.* **2009**, *37*, 1996–2002. [\[CrossRef\]](#)
11. Hanzlíková, I.; Hébert-Losier, K. Is the Landing Error Scoring System reliable and valid? A systematic review. *Sports Health* **2020**. [\[CrossRef\]](#)
12. Padua, D.A.; DiStefano, L.J.; Beutler, A.I.; de la Motte, S.J.; DiStefano, M.J.; Marshall, S.W. The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. *J. Athl. Train.* **2015**, *50*, 589–595. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Dar, G.; Yehiel, A.; Cale'Benzoar, M. Concurrent criterion validity of a novel portable motion analysis system for assessing the landing error scoring system (LESS) test. *Sports Biomech.* **2019**, *18*, 426–436. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Markbreiter, J.G.; Sagon, B.K.; Valovich McLeod, T.C.; Welch, C.E. Reliability of clinician scoring of the landing error scoring system to assess jump-landing movement patterns. *J. Sport Rehabil.* **2015**, *24*, 214–218. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Mauntel, T.C.; Padua, D.A.; Stanley, L.E.; Frank, B.S.; DiStefano, L.J.; Peck, K.Y.; Cameron, K.L.; Marshall, S.W. Automated quantification of the Landing Error Scoring System with a markerless motion-capture system. *J. Athl. Train.* **2017**, *52*, 1002–1009. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv* **2017**, arXiv:1611.08050v2.
17. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [\[CrossRef\]](#)

18. Mathis, M.W.; Mathis, A. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr. Opin. Neurobiol.* **2020**, *60*, 1–11. [\[CrossRef\]](#)
19. Frank, E.; Hall, M.A.; Witten, I.H. The WEKA Workbench. In *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., Eds.; Morgan Kaufmann: Cambridge, MA, USA, 2016.
20. Nadeau, C.; Bengio, Y. Inference for the Generalization Error. *Mach. Learn.* **2003**, *52*, 239–281. [\[CrossRef\]](#)
21. Bland, J.M.; Altman, D.G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **1999**, *8*, 135–160. [\[CrossRef\]](#)
22. Dallinga, J.M.; Benjaminse, A.; Lemmink, K.A.P.M. Which screening tools can predict injury to the lower extremities in team sports? A systematic review. *Sports Med.* **2012**, *42*, 791–815. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Onate, J.; Cortes, N.; Welch, C.; Van Lunen, B.L. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. *J. Sport Rehabil.* **2010**, *19*, 41–56. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.-E.; Sheikh, Y. OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv* **2018**, arXiv:1812.08008v1. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
26. Nan, K.; Liu, S.; Du, J.; Liu, H. Deep model compression for mobile platforms: A survey. *Tsinghua Sci. Technol.* **2019**, *24*, 677–693. [\[CrossRef\]](#)
27. Ekegren, C.L.; Miller, W.C.; Celebrini, R.G.; Eng, J.J.; Macintyre, D.L. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J. Orthop. Sports Phys. Ther.* **2009**, *39*, 665–674. [\[CrossRef\]](#)
28. Whatman, C.; Hing, W.; Hume, P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Phys. Ther. Sport* **2012**, *13*, 87–96. [\[CrossRef\]](#)
29. Willy, R.W. Innovations and pitfalls in the use of wearable devices in the prevention and rehabilitation of running related injuries. *Phys. Ther. Sport* **2018**, *29*, 26–33. [\[CrossRef\]](#)
30. Iqbal, M.H.; Aydin, A.; Brunckhorst, O.; Dasgupta, P.; Ahmed, K. A review of wearable technology in medicine. *J. R. Soc. Med.* **2016**, *109*, 372–380. [\[CrossRef\]](#)
31. Whelan, D.F.; O'Reilly, M.A.; Ward, T.E.; Delahunt, E.; Caulfield, B. Technology in rehabilitation: Evaluating the single leg squat exercise with wearable inertial measurement units. *Methods Inf. Med.* **2017**, *56*, 88–94. [\[CrossRef\]](#)
32. McLean, S.G.; Walker, K.; Ford, K.R.; Myer, G.D.; Hewett, T.E.; van den Bogert, A.J. Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *Br. J. Sports Med.* **2005**, *39*, 355–362. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Willson, J.D.; Davis, I.S. Utility of the frontal plane projection angle in females with patellofemoral pain. *J. Orthop. Sports Phys. Ther.* **2008**, *38*, 606–615. [\[CrossRef\]](#) [\[PubMed\]](#)
34. De Oliveira, F.C.L.; Fredette, A.; Echeverria, S.O.; Batcho, C.S.; Roy, J.S. Validity and reliability of 2-Dimensional video-based assessment to analyze foot strike pattern and step rate during running: A systematic review. *Sports Health* **2019**, *11*, 409–415. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Dos'Santos, T.; McBurnie, A.; Donelon, T.; Thomas, C.; Comfort, P.; Jones, P.A. A qualitative screening tool to identify athletes with 'high-risk' movement mechanics during cutting: The cutting movement assessment score (CMAS). *Phys. Ther. Sport* **2019**, *38*, 152–161. [\[CrossRef\]](#)
36. Myer, G.D.; Ford, K.R.; Hewett, T.E. Tuck jump assessment for reducing Anterior Cruciate Ligament injury risk. *Athl. Ther. Today* **2008**, *13*, 39–44. [\[CrossRef\]](#)

